USGS EROS Download Web Service

1. Introduction

The USGS Seamless Server functionality has recently been expanded to provide a set of webservices to the developer that can be incorporated into custom applications. This document discusses how to use those webservices to request and download data from the USGS Seamless Server data holdings. Three web services will be discussed in this document.

Inventory Service. This is a data discovery service. This service will provide information about what data is available over a particular area of interest. The addition of new datasets and demotion of older datasets from on-line systems to near-line systems is an ongoing process. Updates to the Inventory Service occur on a bi-weekly basis. Efforts are currently underway to include in the Inventory Service all of the datasets currently available through the USGS Seamless Server.

Request Validation Service. Utilizing dataset information obtained from the previous call to the Inventory Service and a user-defined area of interest, this service verifies and validates the information, and then returns to the user fully parameterized URL(s) that can be used by the Download Service.

Download Service. This service initiates a request for data, queries the system to obtain a job status, and returns the requested data to the user.

Orthoimagery data that is no longer considered the "best available" is removed from on-line systems and mapservices on a periodic basis. When this occurs, the dataset is clipped into pre-packaged zipfiles and stored on near-line systems so that it can still be obtained by the public. These datasets currently show up in the Inventory Service with a STATUS = Tiled. We are currently still in the process of creating all of the necessary information that is required to obtain this "historical" data through the Download Service, so a particular tiled dataset may not be available until after December 1, 2009.

2. Download Service

The URL to the Download Service wsdl page is:

http://extract.cr.usgs.gov/axis2/services/DownloadService?wsdl

There are four methods that must be called in the correct sequence:

2.1 initiateDownload

As you can see in the wsdl file, there are over 40 potential parameters that can be passed to this method. This is why it's best to use the Request Validation Service to compose this URL for you. The URL obtained from the Request Validation Service is guraranteed to correctly pass the initiateDownload requirements. This method call will submit the request to the job queue.

The response returned from initiateDownload method call will include a unique id that will be used as input for all successive calls to the Download Service. If there is an error when using this method, a sting containing the word "error" will be returned. The response will look something like this if the call was successful:

<ns:initiateDownloadResponse>
<ns:return>**20090608.150009516.152061160045**</ns:return>
</ns:initiateDownloadResponse>

Your download request is assigned a priority value based on the number of requests you currently have in the job queue. If you only have one request in the queue, it will always be assigned the highest or "best" priority. All jobs of the same priority then are worked on via a First In, First Out basis. If you already have one unfinished job in the queue, then your second job is assigned a priority of one less. Your third unfinished job is assigned a priority of two less than the "best" priority. This is so that one single user cannot use all of the resources when others are also contending for those resources. Therefore, you might want to write your application so that only a certain number of your jobs are in the queue at any time or you might want to monitor the time it takes to work off your jobs and submit new jobs accordingly.

2.2 getDownloadStatus

This method is used to obtain the current status of your request. You must use the job identifier that was returned from initiateDownload method call.

http://extract.cr.usgs.gov/axis2/services/DownloadService/getDownloadStatus?downloadID=20090608.150009516.152061160045

We recommend calling this method no more often than every 30 seconds so that unnecessary status queries don't bog down the servers. As the request passes through

different stages of processing, the status code and text message that is returned will indicate the current progress if you want to pass this on to the requestor.

<ns:return>**210,Extracting Data**</ns:return>

If an error has occurred during one of the processing stages, you will receive a string containing the text "error".

This method is simply checking the status field from the job record that was submitted in the initiateDownload method call. One hour after the job record was first submitted, it is automatically deleted from the database, even if your job has not yet finished processing. Therefore, you may want to stagger your requests and/or submit your jobs during off-hours. If your job has returned an error, try it again. Repeated errors to the same product could indicate a problem with the dataset at EROS and customer services should be notified so that someone at EROS can investigate the issue. As demand for the services increase and resource contention becomes an issue, our architecture is designed such that we can easily and quickly add or reapportion hardware (if available) to process more requests.

2.3 getData

Once you receive back a status of 400 or greater, the finished download bundle has been placed on a webserver and is ready to be downloaded by the user. The download bundle can then be retrieved using the getData method call and the job identifier:

http://extract.cr.usgs.gov/axis2/services/DownloadService/getData?downloadID=20090608.150009516.152061160045

Note that the finished download bundle will be available on the webserver for one hour after the download bundle has been placed there. At that time, an automated process will delete the download bundle to make room for other users. If your application misses the one hour window, then the request must be re-submitted starting at the initiateDownload step.

Depending on what programming language you are using, your application will receive the data as a series of bytes through some kind of connection object. If you have asked for a zip file in your request, you will write the bytes to a file called xxxxxxxx.zip. If you have requested a tar-gzip file as the bundle format, then you will write the bytes to a file called xxxxxxxx.tgz. The data and all support files are contained within the .zip file or the .tgz file. You are then responsible for extracting the information you want from the download bundle.

2.4 setDownloadComplete

Once the download bundle has finished downloading to your computer, please send a download complete message back to the Download Service so that the job is

properly logged and the cleanup process on our webservers can begin. You can do this by using the setDownloadComplete method call and the job identifier:

http://extract.cr.usgs.gov/axis2/services/DownloadService/setDownloadComplete?downloadID=20090608.150009516.152061160045

A successful response returned will look like:

<ns:return>**Successfully updated 20090608.150009516.152061160045 to Complete**</ns:return>